BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 4: e59199
doi: 10.3897/biss.4.59199

OPEN ACCESS

Conference Abstract

Liberating Biodiversity Data From COVID-19 Lockdown: Toward a knowledge hub for mammal host-virus information

Nathan S Upham‡, Donat Agosti§, Jorrit Poelen|, Lyubomir Penev¶, Deborah Paul#, DeeAnn Marie Reeder¤, Nancy B. Simmons«, Gabor Csorba», Quentin Groom˄, Mariya Dimitrova¶, Joseph T Miller˅

‡ Arizona State University, Tempe, United States of America
§ Plazi, Bern, Switzerland
| Ronin Institute, Berkeley, United States of America
¶ Pensoft Publishers & Bulgarian Academy of Sciences, Sofia, Bulgaria
# Florida State University, Tallahassee, United States of America
¤ Bucknell University, Lewisburg, United States of America
« American Museum of Natural History, New York, United States of America
» Hungarian Natural History Museum, Budapest, Hungary
˄ Meise Botanic Garden, Meise, Belgium
˅ Global Biodiversity Information Facility, Copenhagen, Denmark

Corresponding author: Nathan S Upham (nathan.upham@asu.edu)

Received: 01 Oct 2020 | Published: 09 Oct 2020

Citation: Upham NS, Agosti D, Poelen J, Penev L, Paul D, Reeder DM, Simmons NB, Csorba G, Groom Q, Dimitrova M, Miller JT (2020) Liberating Biodiversity Data From COVID-19 Lockdown: Toward a knowledge hub for mammal host-virus information. Biodiversity Information Science and Standards 4: e59199. https://doi.org/10.3897/biss.4.59199

Abstract

A deep irony of COVID-19 likely originating from a bat-borne coronavirus (Boni et al. 2020) is that the global lockdown to quell the pandemic also locked up physical access to much basic knowledge regarding bat biology. Digital access to data on the ecology, geography, and taxonomy of potential viral reservoirs, from Southeast Asian horseshoe bats and pangolins to North American deer mice, was suddenly critical for understanding the disease's emergence and spread. However, much of this information lay inside rare books and personal files rather than as open, linked, and queryable resources on the internet. Even the world's experts on mammal taxonomy and zoonotic disease could not retrieve their data from shuttered laboratories. We were caught unprepared. Why, in this digitally

© Upham N et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

connected age, were such fundamental data describing life on Earth not already freely accessible online?

Understanding why biodiversity science was unprepared—and how to fix it before the next pandemic—has been the focus of our COVID-19 Taskforce since April 2020 and is continuing (organized by CETAF and DiSSCo). We are a group of museum-based and academic scientists with the goal of opening the rich ecological data stored in natural history collections to the research public. This information is rooted in what may seem an unlikely location—taxonomic names and their historical usages, which are the keys for searching literature and extracting linked ecological data (Fig. 1). This has been the core motivation of our group, enabled by the pioneering efforts of Plazi (Agosti and Egloff 2009) to build tools for literature digitization, extraction, and parsing (e.g., Synospecies, Ocellus) without which biodiversity science would be even less prepared. Our group led efforts to build an additional pipeline from Plazi to the Biodiversity Literature Repository at Zenodo, a free and unlimited data repository (Agosti et al. 2019), and then to GloBI, an open-source database of biotic interactions (Poelen et al. 2014, GloBI 2020). We also developed a direct integration from Pensoft Journals to GloBI, leveraging that publisher's indexing of computer-readable terms (called semantic metadata; Senderov et al. 2018) to extract mammal host and virus information.

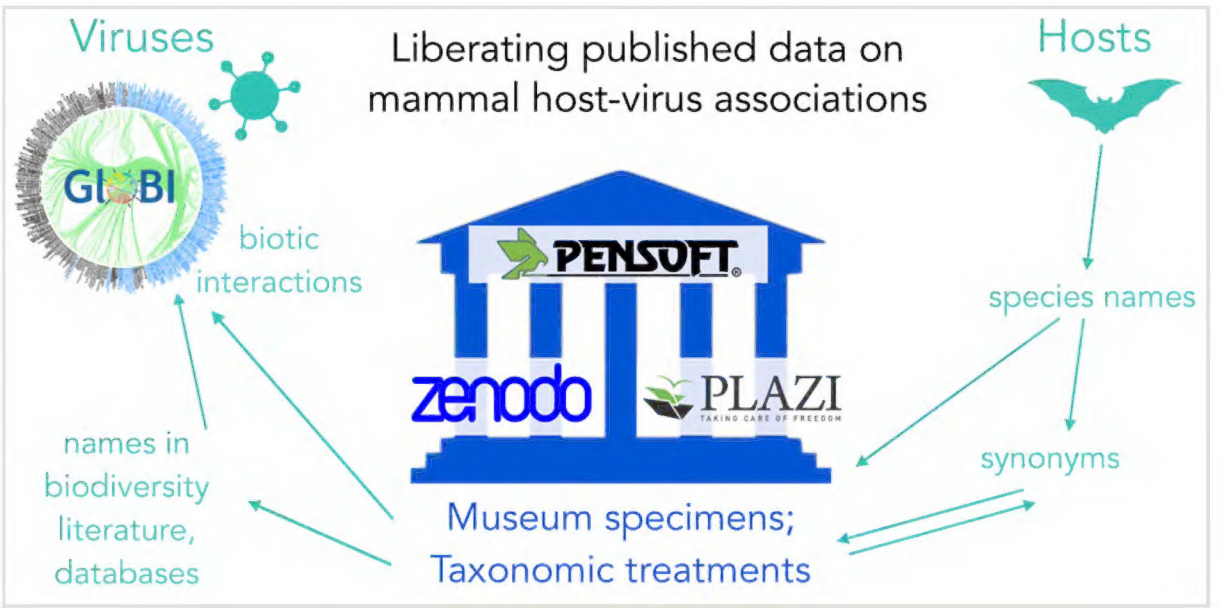

Figure 1.

Taxonomic names and their usages are the key for unlocking host-virus interaction data. Flow of information from digitizing taxonomic treatments containing species names and their historical usages (Plazi and Zenodo), to searching biodiversity literature for data linked to names, to connecting those biotic interactions in an ecological network (GloBI). Data can also flow directly into GloBI from Pensoft-style journals that publish data with computer-readable tags.

Overall, considerable progress was made. In total, 85,492 new interactions were added to GloBI from 14 April to 21 May 2020 (see entire dataset on Zenodo: Poelen et al. 2020). Of those, 28,839 interactions are present when subset to "hasHost", "hostOf", "pathogenOf", "virus", and 4,101 unique name combinations are present after considering mammal species synonymies (from Meyer et al. 2015). Of those interactions, 892 species of

mammals and 1,530 unique virus names are involved, which compares to 754 mammals and 586 viruses in the most recent data synthesis (Olival et al. 2017). While these liberated data may still include redundancies, they demonstrate the value of our approach and the expanse of known but digitally unconnected data that remains locked in publications.

We can liberate host-virus data from publications, but doing so is expensive and does not scale to the continued influx of new articles that are inadequately digitized. Our efforts make it clear that Pensoft-style semantic publishing should be expanded to all major journals. The pandemic has created an opportunity for re-thinking the way we do science in the digital age. Thankfully, our future is not the past, so we do not have to keep wasting resources to digitially 'rediscover' biodiversity knowledge. We collectively call for changes to the publishing paradigm, so that research findings are directly accessible, citable, discoverable, and reusable for creating complete forms of digital knowledge.

Keywords

zoonotic disease risk, spillover, virus, mammal, bat, taxonomy, semantic publishing

Presenting author

Nathan S Upham

Presented at

TDWG 2020

References

- Agosti D, Egloff W (2009) Taxonomic information exchange and copyright: the Plazi approach. BMC Research Notes 2 (1). https://doi.org/10.1186/1756-0500-2-53
- Agosti D, Catapano T, Sautter G, Kishor P, Nielsen L, Ioannidis-Pantopikos A, Bigarella C, Georgiev T, Georgiev T, Penev L, Egloff W (2019) Biodiversity Literature Repository (BLR), a repository for FAIR data and publications. Biodiversity Information Science and Standards 3 https://doi.org/10.3897/biss.3.37197
- Boni M, Lemey P, Jiang X, Lam TT, Perry B, Castoe T, Rambaut A, Robertson D (2020) Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nature Microbiology 1-10. https://doi.org/10.1038/s41564-020-0771-4
- GloBI (2020) Plazi-Zenodo-GloBI integration. URL: https://www.globalbioticinteractions.org/plazi-zenodo/
- Meyer C, Kreft H, Guralnick R, Jetz W (2015) Global priorities for an effective information basis of biodiversity distributions. Nature Communications 6 https://doi.org/10.1038/ncomms9221

- Olival K, Hosseini P, Zambrana-Torrelio C, Ross N, Bogich T, Daszak P (2017) Host and viral traits predict zoonotic spillover from mammals. Nature 546 (7660): 646-650. https://doi.org/10.1038/nature22975
- Poelen J, Simons J, Mungall C (2014) Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets. Ecological Informatics 24: 148-159. https://doi.org/10.1016/j.ecoinf.2014.08.005
- Poelen J, Upham N, Agosti D, Ruschel T, Guidoti M, Reeder D, Simmons N, Penev L, Dimitrova M, Csorba G, Groom Q, Willoughby A (2020) CETAF-DiSCCo/COVID19-TAF biodiversity-related knowledge hub working group: indexed biotic interactions and review summary. Zenodo. DOI: 10.5281/zenodo.3839098 type: dataset. URL: https://zenodo.org/record/3839098#.XtD-C8YpBTY
- Senderov V, Simov K, Franz N, Stoev P, Catapano T, Agosti D, Sautter G, Morris R, Penev L (2018) OpenBiodiv-O: ontology of the OpenBiodiv knowledge management system. Journal of Biomedical Semantics 9 (1). https://doi.org/10.1186/s13326-017-0174-5